



# Forschungsdateninfrastrukturen und deren Bedeutung für moderne Forschung

Prof. Dr. Sonja Schimmler

07.12.2023, KI in Bibliotheken - Neue Wege mit großen Sprachmodellen?

# **Digitalisation and Opening Science Research Data Infrastructures**

## Digitalisation and Opening Science

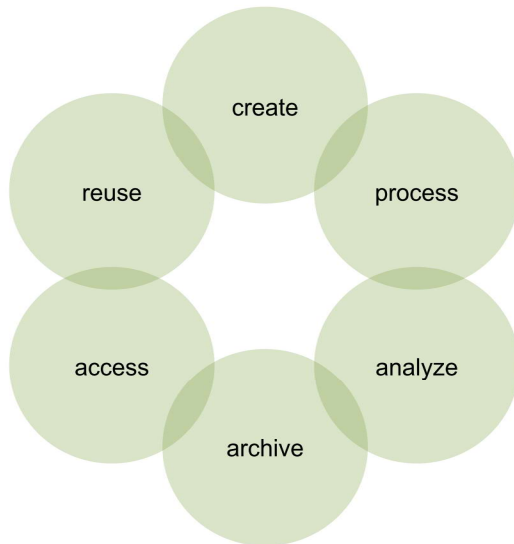
- Modern research relies on more and more data
- Modern research on an internationally competitive level is no longer possible without digital support



- This leads to a paradigm shift
- Scientific progress will only be furthered, if
  - (1) data becomes available at large scale
  - (2) data is linked and machine-interpretable
  - (3) specificities of the different disciplines are very carefully taken into account

# Research Data Infrastructures

- **Considering the whole research data lifecycle**



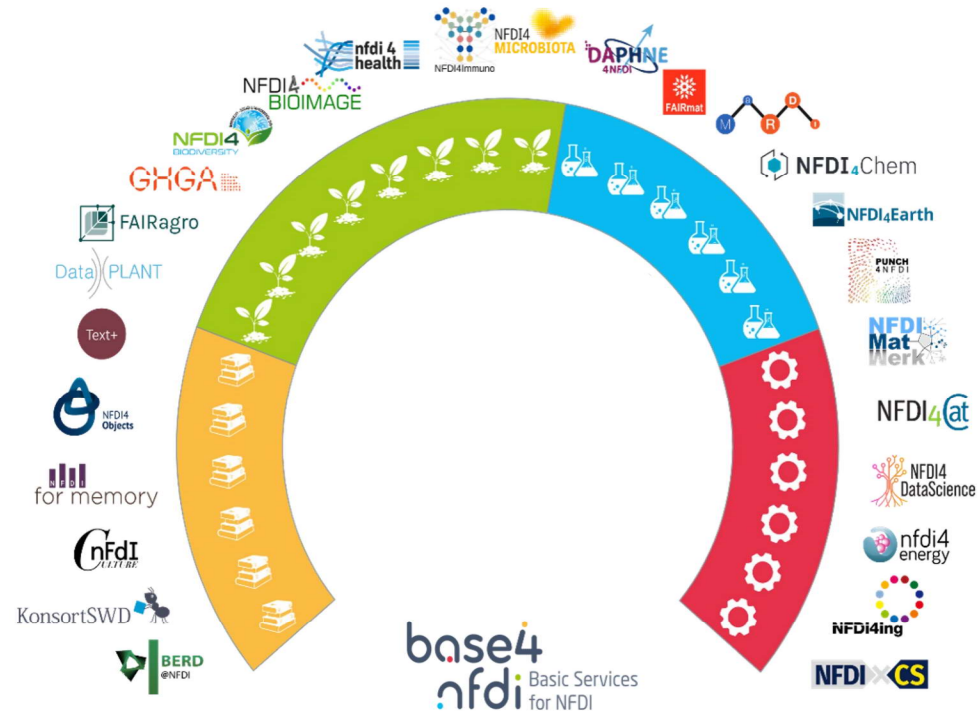
- Besides publications and research data, other **digital artefacts** such as research software should be taken into account and should be properly represented
- Digital artefacts should be **enriched with metadata** and should be **linked** for contextualization
- This forms the technical foundation for **modern research data infrastructures** and **tools**
- Special emphasis should be put on the **usability** of the resulting systems

**NFDI**

# National Research Data Infrastructure (NFDI)



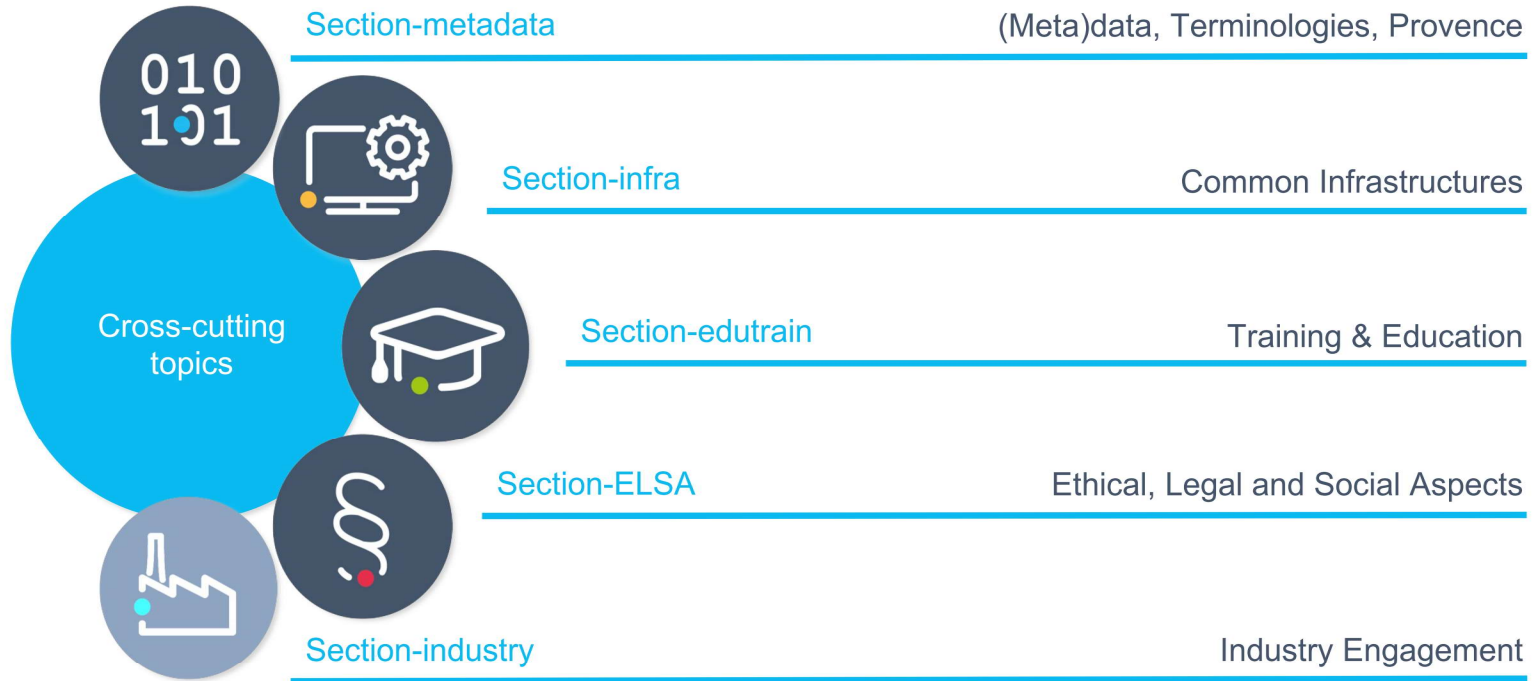
- NFDI to set up a **German National Research Data Infrastructure**
  - The goal is to build **ONE NFDI**
- 26 NFDI consortia funded



# National Research Data Infrastructure (NFDI)



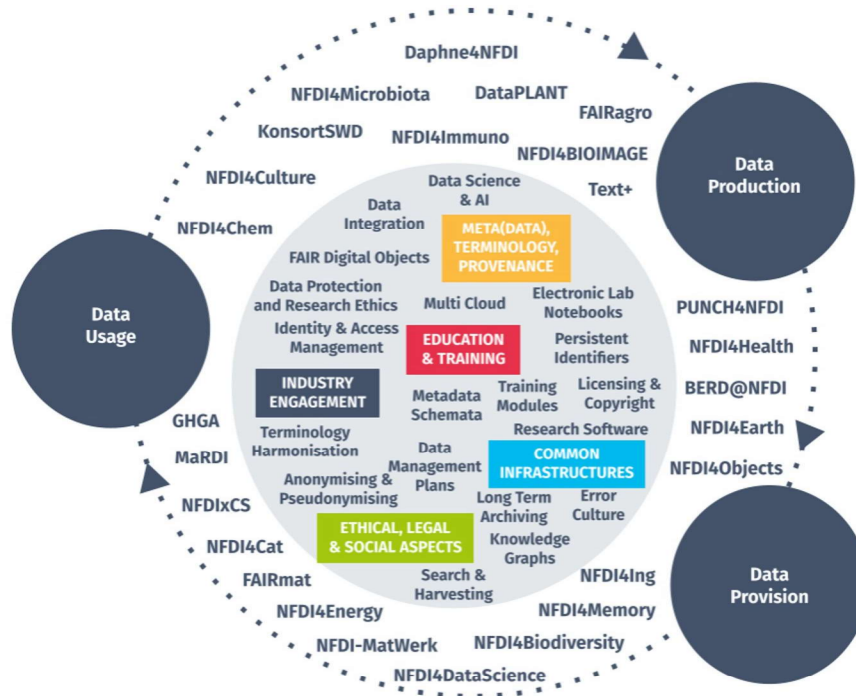
- Sections to work on cross-cutting topics



# National Research Data Infrastructure (NFDI)



- Base4NFDI to develop basic services



## IAM4NFDI

**Identity & Access Management**  
Supported by: Section Common Infrastructures

## PID4NFDI

**Persistent Identifier Service**  
Supported by: Section Common Infrastructures

## TS4NFDI

**Terminology Services**  
Supported by: Section Metadata, Terminologies, Provenance



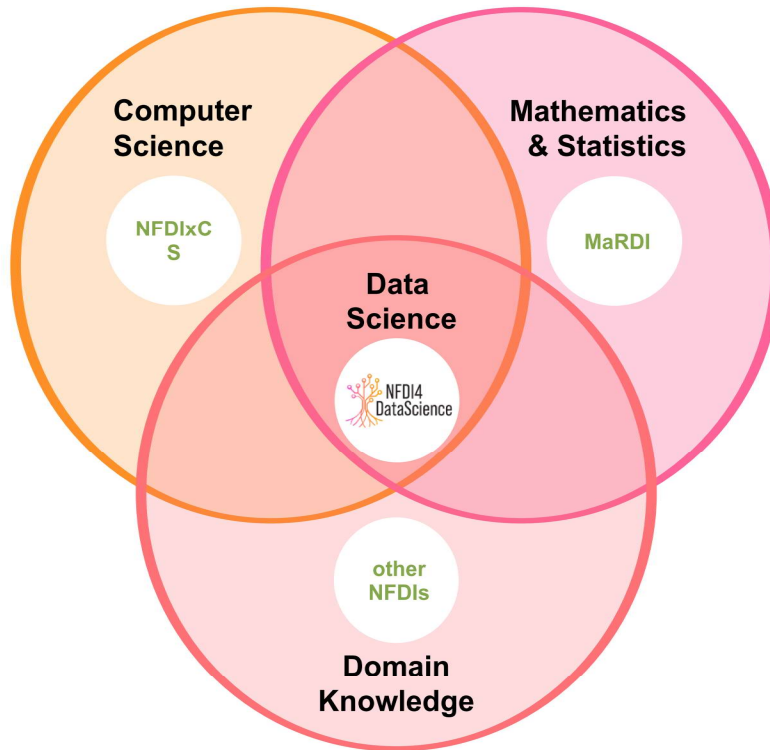
**NFDI4DataScience**

# NFDI4DS: NFDI for Data Science and Artificial Intelligence



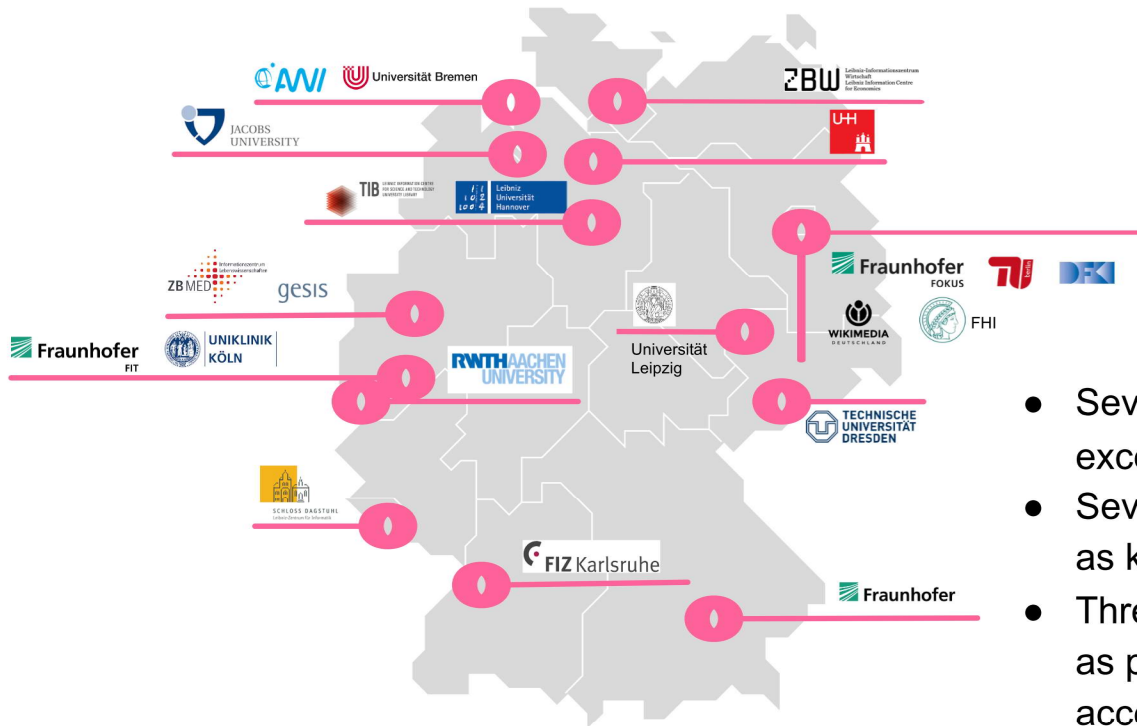
- NFDI4DS to set up a
  - **National Research Data Infrastructure for Data Science and Artificial Intelligence**
- **Our Vision**
  - Target the **data science** and **artificial intelligence community** in Germany
  - Support the **whole research data lifecycle**: collect/create, process, analyze, preserve, access, reuse
- **Our Objectives**
  - Make **all digital objects** (articles, data, models, workflows, scripts/code) available
  - **Interlink** all digital objects and offer innovative tools and services
  - Develop, establish and sustain a national **research data infrastructure**
  - Foster **FAIR** (findable, accessible, interoperable, reusable) and open research
  - Increase **transparency, reproducibility** and **fairness** in research
  - Consider **ethical, legal** and **social** aspects in research

## NFDI4DS: Context



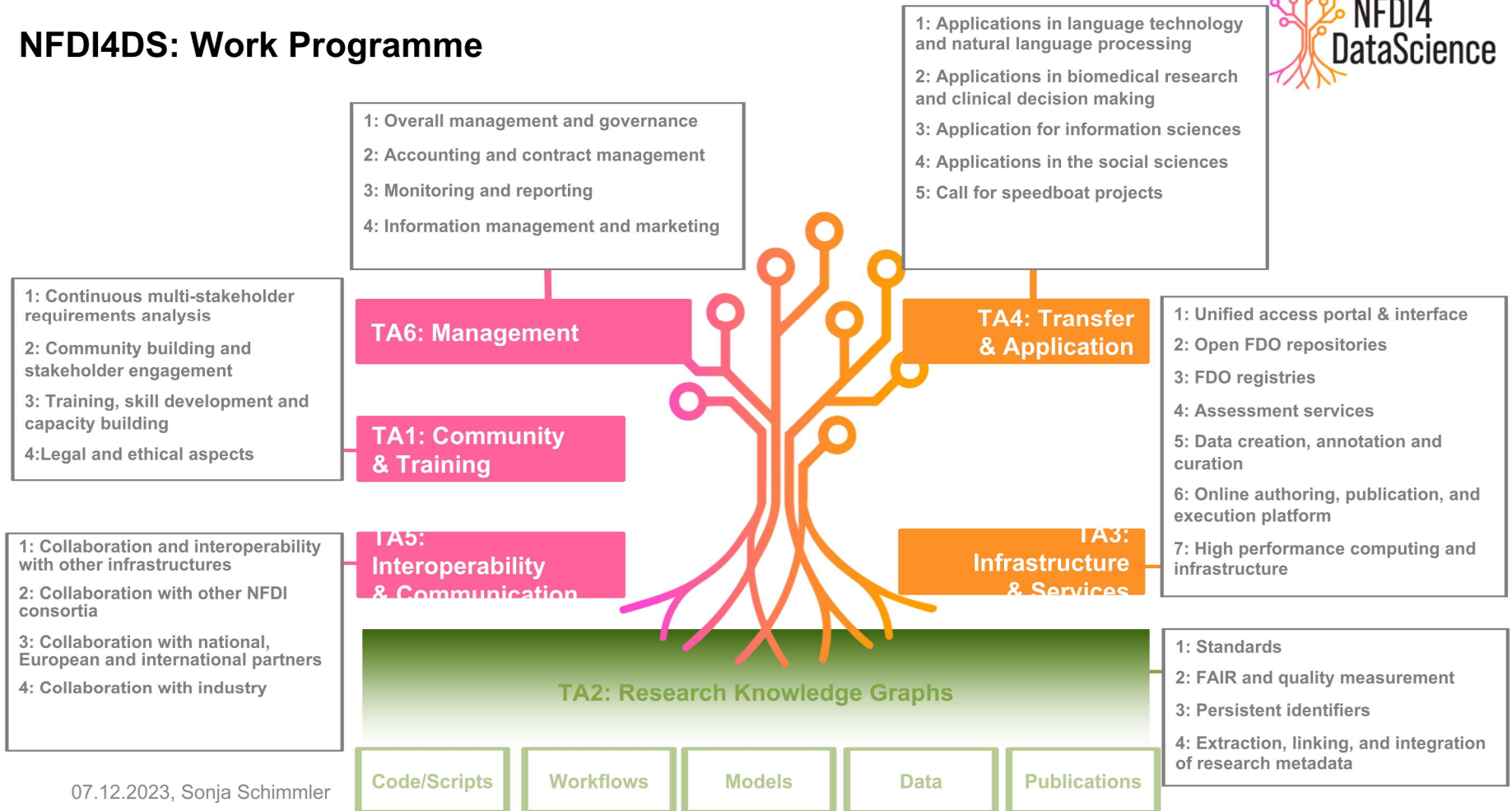
- Methods from **computer science** form the prior foundation.
- Solutions from **mathematics and statistics** are key for new models and evaluations.
- **Domain expertise** from different domains and from a wide range is the driver to bring solutions into all domains alike.

# NFDI4DS: Consortium



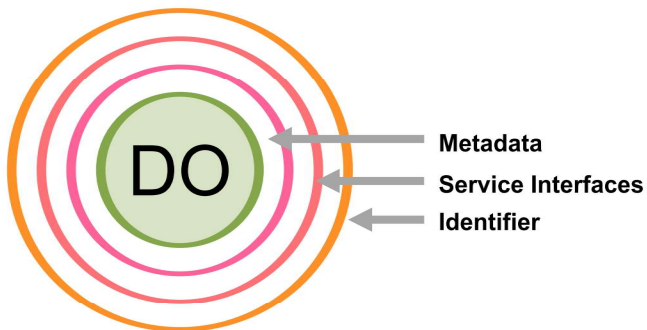
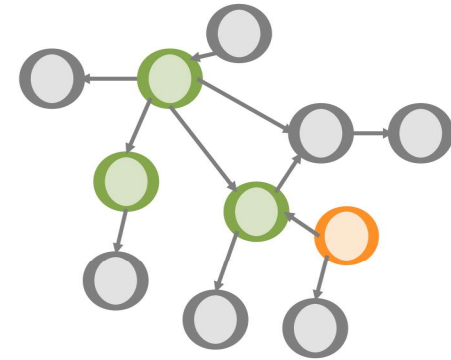
- Several **universities** as centers of excellence in DS and AI.
- Several **non-university research institutes** as key contributors.
- Three large scientific **information centers** as providers and maintainers of data and for access to domain user groups.

# NFDI4DS: Work Programme



## NFDI4DS: Knowledge Graphs (KG)

- **Open knowledge base** of DS & AI resource (code, data and models) **metadata & their relations**.
- State-of-the-art **data representation** to implement FAIR principles.
- **Open & transparent resource** for searching research metadata and as **public resource/infrastructure** through open APIs.



## NFDI4DS: FAIR Digital Objects (FDO)

- **Stable actionable unit** enabling reliable interpretation and processing of the contained metadata and resources (code, data and models).
- **Representation** of a wide **range of resources** based on FAIR principles.

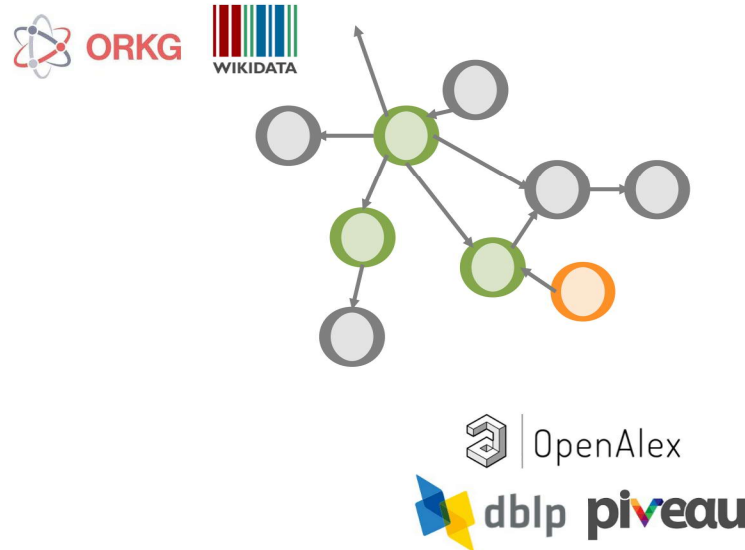
# NFDI4DS: FAIR Digital Objects & Research Knowledge Graphs



**Publications**

**Data**

## Knowledge Graphs - *Semantic Description*

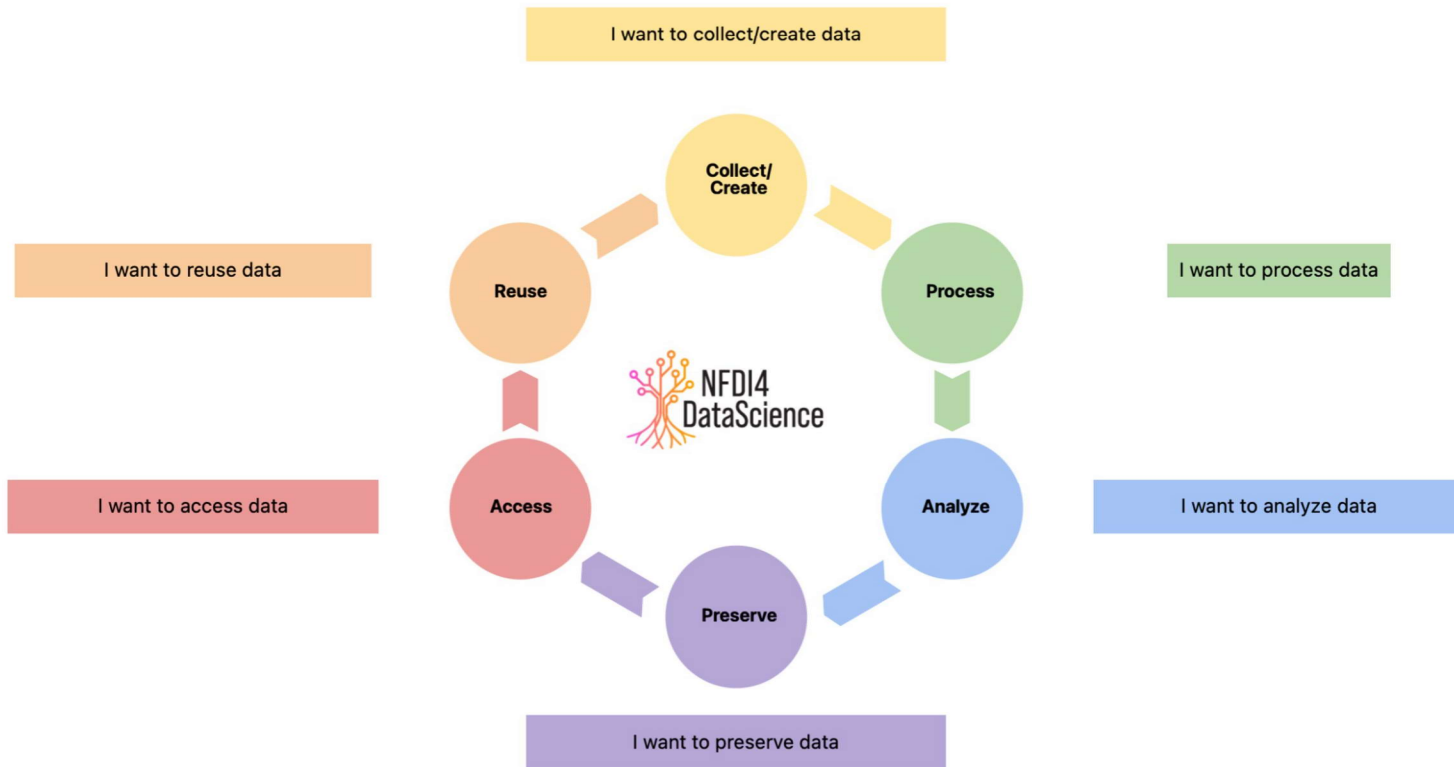


**Code / Scripts**

**Models**

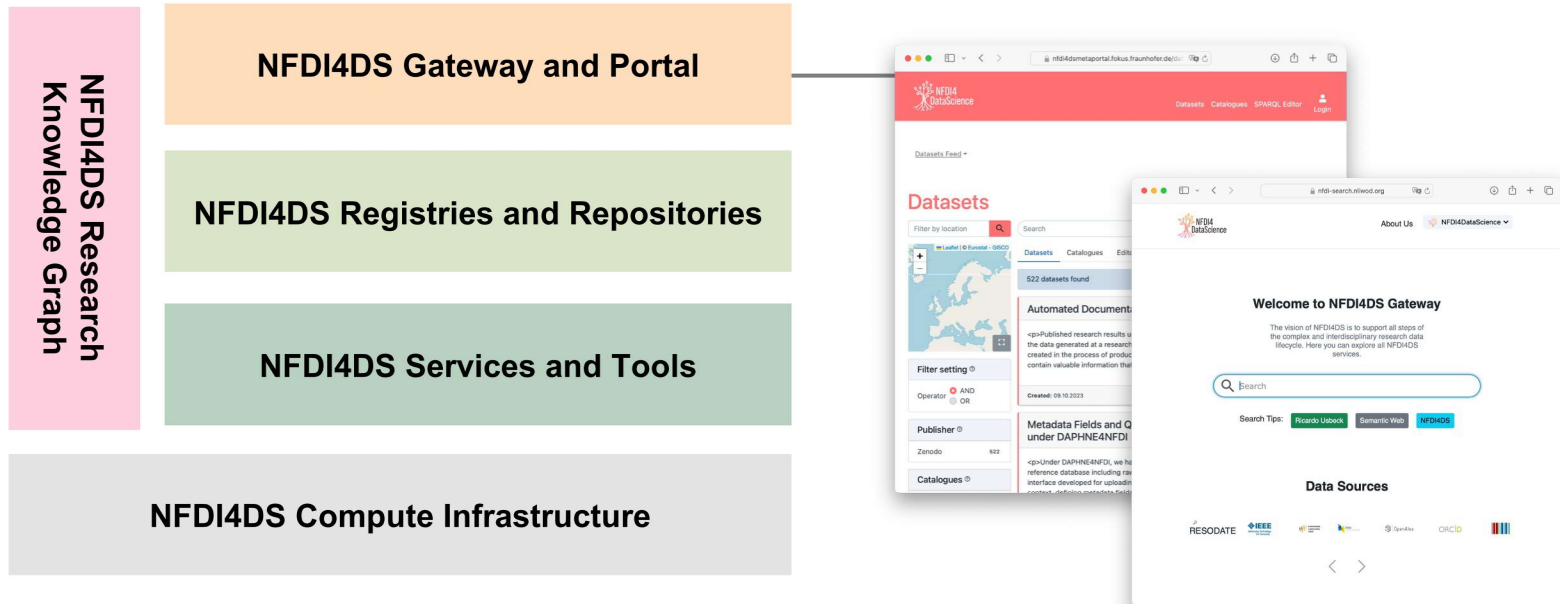
**Workflows**

# NFDI4DS: Research Data Lifecycle





# NFDI4DS: Infrastructure and Services



*What needs to be adapted to support researchers that use LLMs?*

*Can LLMs form the basis for new infrastructure and service components?*

## NFDI4DS: Community Involvement - New Technologies


Language Technology	DFKI
Life Sciences	ZB MED
Information Sciences	TIB
Social Sciences	GESIS

*Is there a difference in the use of LLMs in the different communities?*

- **Speed Boat Projects** to start with other communities


# NFDI4DS: Community Involvement - New Technologies



 The Open Modeling Foundation:  
A Global Community for  
Standards-Based Modeling  
of Human and Natural Systems

21<sup>st</sup> June 2023 5:00 p.m. online

Lecture Series NFDI4DS



www.nfdi4datascience.de


 Beatriz Serrano-Solano  
Introduction to AI4Life

24. Oct 2023 9:00 online

Lecture Series NFDI4DS




www.nfdi4datascience.de

 Introduction to NFDI4DS and  
OpenCitations

3<sup>rd</sup> April 2023 10:00 a.m. online

Lecture Series NFDI4DS



www.nfdi4datascience.de

 Marco Jahn, Eclipse Foundation  
Software ate the world - and Open  
Source is eating software!

28. Nov 2023 9:00 online

Lecture Series NFDI4DS



www.nfdi4datascience.de

## NFDI4DS Lecture Series

*How can we support researchers  
and the general public  
understanding new technologies?*

**NFDI4DS Conference  
& Consortium Meeting**

9<sup>th</sup> Nov 2023 Conference  
10<sup>th</sup> Nov 2023 Consortium Meeting



www.nfdi4datascience.de

## NFDI Conference & Consortium Meeting (2021, 2022, 2023)

**NFDI Science Slam**

9<sup>th</sup> Nov 2023 6:00 p.m.  
Location: Fraunhofer FOKUS, Berlin

In Cooperation with Berlin Science Week



www.nfdi4datascience.de

## NFDI Science Slam (2021, 2022, 2023)

# NFDI4DS: Community Involvement - New Technologies



weizenbaum  
foru



**Weizenbaum-Forum: ChatGPT, Stable Diffusion & Co.**  
Welche Auswirkungen haben so genannte generative KI-Modelle auf die Gesellschaft?

9. Mai 2023 | 18:00-19:30 Uhr | digitale Veranstaltung

Forum: ChatGPT, Stable Diffusion & Co

*How can we support researchers  
and the general public  
understanding new technologies?*

weizenbaum  
foru

„ Unsere öffentlichen Diskurse über KI brauchen einen Perspektivwechsel – hin zu ihren grundlegenden Funktionsweisen, den damit verbundenen Möglichkeiten und Grenzen. Dadurch rücken wichtigere Fragen in den Fokus als die, ob uns KI ersetzen wird. Wir sollten uns den Gestaltungsspielräumen zuwenden, anstatt Angst zu haben oder einfach einem Hype zu folgen.

Joana Müller-Brehe  
Wissenschaftliche Mitarbeiterin  
im Zentrum für vertrauenswürdiges KI (ZVKI)  
und Data Analytics Lab



weizenbaum  
foru

„ Als Kunstvermittlerin interessieren mich besonders die Machtverhältnisse, Bildlogiken und Wissensproduktionen künstlicher Intelligenzen. Ich sehe es als essentielle Aufgabe, dies kreativ, kritisch und künstlerisch zu hinterfragen sowie mitzugestalten!

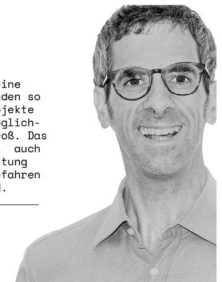
Anja Gebauer  
Research Associate Digital Education  
Hamburger Kunsthalle



weizenbaum  
foru

„ Ich habe es selten erlebt, daß eine neue Technologie bei unseren Kunden so schnell Einfluss auf konkrete Projekte genommen hat. Die Chancen und Möglichkeiten für neue Produkte sind groß. Das ist sehr spannend, aber zum Teil auch gefährlich, da die Erwartungshaltung hoch ist und Limitationen und Gefahren nicht gleich offensichtlich sind.

Hans-Peter Zann  
Head of Artificial Intelligence / CTO



## NFDI4DS: Community Involvement - New Technologies

- **Shared Tasks in 2023/2024**

- Scholarly QALD @ **ISWC 2023**
  - Question answering over scholarly knowledge graphs
- SOTA @ **CLEF 2024**
  - State-of-the-art tracking in AI scholarly publications
- FoRC @ **ESWC 2024**
  - Field of Research Classification for Scholarly Publications
- SOMD @ **ESWC 2024**
  - Software Mention Detection in Scholarly Publications

- **Mini Hackathons in 2023**

- Metadata for machine learning
- Machine-actionable software management plans
- FAIRification game for software

*Can LLMs form the basis for new infrastructure and service components?*

# NFDI4DS: Community Involvement - New Technologies



## User Interviews: Ethical, Legal and Social Aspects

Ethical, legal and social implications of Data Science and AI are turning out to be an important issue that is increasingly relevant for data scientists, policymakers and the general public.

NFDI4DataScience plans to identify these challenges and issue recommendations on how they should be addressed. In order to achieve this goal we are gathering requirements via interviews with researchers and practitioners that are/were involved in Data Science and AI projects.

Are you a Data Scientist? Are you concerned with the sometimes disparate impact of your work? Are you feeling lost in the jungle of numerous legal provisions? Then you can help us by describing the specific ethical, legal and social challenges you have faced in any of your projects. An interview will take about 60 minutes and will be conducted online.

## NFDI4DS Interviews: Ethical, Legal & Social Aspects

*How can we support researchers and the general public understanding new technologies?*



NFDI4DS Video Series: Conversations on AI Ethics



# Forschungsdateninfrastrukturen und deren Bedeutung für moderne Forschung

Prof. Dr. Sonja Schimmler

07.12.2023, KI in Bibliotheken - Neue Wege mit großen Sprachmodellen?